

Optimal transportation with infinitely many marginals*

Brendan Pass[†]

March 8, 2013

Abstract

We formulate and study an optimal transportation problem with infinitely many marginals; this is a natural extension of the multi-marginal problem studied by Gangbo and Swiech [14]. We prove results on the existence, uniqueness and characterization of the optimizer, which are natural extensions of the results in [14]. The proof relies on a relationship between this problem and the problem of finding barycenters in the Wasserstein space, a connection first observed for finitely many marginals by Agueh and Carlier [1].

1 Introduction

In this paper, we study an optimal transportation problem with infinitely many marginals.

Optimal transportation with two marginals is an exciting and fast moving area of research. The general goal is to couple two probability measures together as efficiently as possible, relative to a given cost function. More precisely, given measures μ_1 and μ_2 (called *marginals*) on topological spaces M_1 and M_2 , respectively, and a cost function $c : M_1 \times M_2 \rightarrow \mathbb{R}$, the aim is to find the measure γ on $M_1 \times M_2$ which projects to μ_1 and μ_2 and minimizes the total transportation cost:

$$\int_{M_1 \times M_2} c(x_1, x_2) d\gamma$$

Equivalently, one can formulate this problem using more probabilistic language. Here one looks for an $M_1 \times M_2$ valued random variable (X_1, X_2) , such that $\text{law} X_i = \mu_i$, for $i = 1, 2$, which minimizes the expectation:

$$E[c(X_1, X_2)]$$

*The author was supported in part by a University of Alberta start-up grant.

[†]Department of Mathematical and Statistical Sciences, 632 CAB, University of Alberta, Edmonton, Alberta, Canada, T6G 2G1 pass@ualberta.ca.

Results about the existence, uniqueness and structure of the optimal measure γ have been proven for a wide class of cost functions and marginals; for a detailed review, see the monograph of Villani [29]. A central theme is that, under certain conditions on the cost and the measures, there is a unique optimal measure γ , concentrated on the graph of a function, $x_2 = F(x_1)$; this was first proven for the quadratic cost $c(x_1, x_2) = |x_1 - x_2|^2$ on $M_1 = M_2 = \mathbb{R}^n$ by Brenier [4] and was generalized to a large class of cost functions by Gangbo [12], Gangbo and McCann [13], Caffarelli [5], McCann [20] and Levin [18]. In probabilistic terms, this means that the random variables (X_1, X_2) are completely dependent.

In recent years, optimal transportation problems with several marginals have started to attract more attention; this is a natural generalization of the preceding problem. Give m probability measures $\mu_1, \mu_2, \dots, \mu_m$ on topological spaces M_1, M_2, \dots, M_m , and a cost function $c : M_1 \times M_2 \times \dots \times M_m \rightarrow \mathbb{R}$, we look for the measure γ on the product $M_1 \times M_2 \times \dots \times M_m$ which projects to the μ_i respectively, and minimizes

$$\int_{M_1 \times M_2 \times \dots \times M_m} c(x_1, x_2, \dots, x_m) d\gamma$$

As in the two marginal case, this problem may be formulated probabilistically. In this setting, one looks for an $M_1 \times M_2 \times \dots \times M_m$ valued random variable (X_1, X_2, \dots, X_m) , such that $\text{law}(X_i) = \mu_i$ for $i = 1, 2, \dots, m$, minimizing

$$E[c(X_1, X_2, \dots, X_m)]$$

In contrast to the two marginal case, results concerning the structure of the optimal measure for $m > 2$ are rather scarce. However, Gangbo and Swiech proved that for the cost function $c(x_1, x_2, \dots, x_m) = \sum_{i=1}^m \sum_{j=1}^m |x_i - x_j|^2$ on $M_i = \mathbb{R}^n$, the Kantorovich problem admits a unique solution which is concentrated on the graph of a function over the first marginal, generalizing Brenier's theorem [14]; see also [22][17][26] and [27]. As in the two marginal case, this means that the random variables (X_1, X_2, \dots, X_m) are completely dependent. Since then, a handful of results have been proven on the structure of solutions for different cost functions by Heinich [15], Carlier [6], Carlier and Nazaret [8] and the present author [25][24]. Applications for multi-marginal optimal transportation have also arisen in mathematical economics [7][9] and condensed matter physics [11][10].

Our goal in the present article is to study this problem in the limit as $m \rightarrow \infty$, restricting our attention to a cost function reminiscent of that of Gangbo and Swiech. More precisely, we will prescribe a *continuum* of probability measures μ_t on \mathbb{R}^n , for $t \in [0, 1]$. We will then look for the measurable¹ stochastic process, X_t , with single time marginals $\text{law}(X_t) = \mu_t$, that minimizes

¹By definition, the stochastic process $X_t = X_t(\omega)$ is a mapping from $\Omega \times [0, 1] \rightarrow \mathbb{R}^n$, where Ω is a probability space. By measurable, we mean that this mapping is measurable, with respect to product measure on $\Omega \times [0, 1]$; by Fubini's theorem, this implies that the sample paths $t \mapsto X_t$ are measurable almost surely.

$$E\left(\int_0^1 \int_0^1 |X_s - X_t|^2 ds dt\right) \quad (MK_\infty)$$

After expanding $|X_s - X_t|^2$ and noting that, by Fubini's theorem, $E(\int_0^1 X_t^2 dt) = \int_0^1 E(X_t^2) dt = \int_0^1 \int_{\mathbb{R}^n} x^2 d\mu_t(x) dt$, for *any* measurable process such that $\text{law} X_t = \mu_t$ for all t , it is clear that this is equivalent to *maximizing*:

$$E\left(\left(\int_0^1 X_t dt\right)^2\right)$$

We can think of the function $\int_0^1 \int_0^1 |X_s - X_t|^2 ds dt$ as the limit of the Gangbo and Swiech cost. On the other hand, for a sample path X_t , the integral $\int_0^1 X_t dt$ represents the average position of the sample path. If we think of X_t as representing a particle moving in a quadratic potential, over a time period $t \in [0, 1]$, then $(\int_0^1 X_t dt)^2$ is the potential of the average position of the particle.

Our main result, Theorem 4.2.2, asserts existence and uniqueness of an optimizer in (MK_∞) , as well as a characterization of it, and is the natural generalization of the result of Gangbo and Swiech from finitely many to infinitely many marginals. Roughly speaking, it says that the random curve X_t is completely dependent, or deterministic; if X_{t_0} is known for one fixed t_0 , then X_t is known for all t (see Theorem 4.2.4).

The typical approach to optimal transportation problems (with finitely many marginals) is to develop a duality theory, and then to use the resulting first order conditions to derive structural results about the optimal measure. Our strategy here is quite different. A recent paper of Agueh and Carlier relates the multi-marginal problem with Gangbo and Swiech's cost function to barycenters in the Wasserstein space [1]. In this paper, we first generalize their results on existence, uniqueness and regularity from barycenters of finitely many points to barycenters of curves. Having done this, we adapt their relationship between barycenters and multi-marginal problems to our setting and then exploit this connection to deduce the existence and uniqueness of the solution to our problem.

Barycenters in the Wasserstein space are an interesting topic in their own right. Barycenters of probability measures on general length spaces have attracted quite a bit of attention recently, in large part because of their relationship to curvature. In spaces with Alexandrov curvature bounded above, the behaviour of barycenters is fairly well understood; see the work of Sturm [28]. The study of barycenters on spaces with curvature bounded below has recently been initiated by Ohta, and remains in its infancy [21]. It is, however, already apparent that barycenters on spaces with lower curvature bounds are not as well behaved as their counterparts on spaces with upper curvature bounds. In particular, on spaces with non-positive curvature, each measure admits a unique barycenter, whereas on spaces with non-negative curvature, barycenters may be non-unique. As an elementary example, every point on the equator is a barycenter of the north and south pole on the unit sphere.

In addition, the problem of interpolating among several probability measures has begun to arise in applied problems including texture mixing [3] and mathematical economics [7][9]. In fact, in [7], the authors also consider an extension of their model which involves interpolating among an infinite number of measures.

It is well known that the Wasserstein space over \mathbb{R}^n does not have non-positive Alexandrov curvature [2]. The work of Agueh and Carlier provides uniqueness and regularity results, as well as a characterization of the barycenter of finitely many points in the Wasserstein space, under certain regularity conditions. Our first contribution is to extend their uniqueness and regularity result to a continuous curve μ_t of measures. It is also worth noting that our techniques here can be used to extend some of the results of Agueh and Carlier to other underlying spaces; in particular, we prove uniqueness of barycenters in the Wasserstein space over a Riemannian manifold.

Finally, let us mention that in a separate paper, we study infinite marginal optimal transportation for somewhat more general cost functions, restricted to the case $n = 1$ [23]. The techniques used there are quite different than here.

In the next section, we will introduce our hypotheses on the curve of measures, μ_t , as well as two regularity assumptions which will be assumed only at specific points. In the third section we will study the barycenter of the curve μ_t proving existence, uniqueness and regularity, as well as demonstrating that the uniqueness result can be extended to other settings. In section 4, we develop the connection between barycenters and the problem (MK_∞) and use this to prove existence and uniqueness of the optimal stochastic process.

2 Notation and assumptions

We will denote by $P_2(\mathbb{R}^n)$ the set of all probability measures on \mathbb{R}^n with finite second moments and $P_{ac,2}(\mathbb{R}^n)$ the subset of these which are absolutely continuous with respect to Lebesgue measure. For $\mu, \nu \in P_2(\mathbb{R}^n)$, $W_2(\mu, \nu)$ denotes the quadratic Wasserstein distance between the measures μ and ν :

$$W_2^2(\mu, \nu) = \inf \int_{\mathbb{R}^n \times \mathbb{R}^n} |x - y|^2 d\gamma(x, y)$$

where the infimum is taken over all Borel probability measures γ on $\mathbb{R}^n \times \mathbb{R}^n$ projecting to μ and ν , respectively.

Let $M \subseteq \mathbb{R}^n$ be a bounded domain and $P(M) \subseteq P_2(\mathbb{R}^n)$ be the set of all Borel probability measures on M . We will assume that all of our measures μ_t are supported on M ; that is, $\mu_t \in P(M)$. We will denote by $c(M)$ the convex hull of M and $P(c(M))$ the set of Borel probability measures on $c(M)$. We will assume that μ_t is a weakly continuous curve in $P(M)$; that is, we assume the mapping $t \mapsto \mu_t$ is a continuous map with respect to the weak topology. Note that, by the boundedness of M , this is equivalent to continuity with respect to the Wasserstein metric.

We now introduce two different regularity conditions on the μ_t , which we will assume at different times.

Assumption A. *The set*

$$A := \left\{ t : \mu_t = g_t(x)dx \text{ is absolutely continuous with respect to Lebesgue measure.} \right\}$$

has positive Lebesgue measure.

Assumption B. *The set*

$$A_\infty := \left\{ t : \mu_t = g_t(x)dx \text{ is absolutely continuous with respect to Lebesgue measure and } \|g_t\|_{L^\infty} < \infty \right\}$$

has positive Lebesgue measure.

Note that assumption B easily implies that for some $K < \infty$, the set

$$A_K := \left\{ t : \mu_t = g_t(x)dx \text{ is absolutely continuous with respect to Lebesgue measure and } \|g_t\|_{L^\infty} \leq K \right\}$$

has positive Lebesgue measure.

Finally, note that Assumption B clearly implies Assumption A.

3 Barycenters

In this section, we study the barycenter of the measures μ_t . By definition, this is the minimizer of

$$\mu \mapsto \int_0^1 W_2^2(\mu, \mu_t) dt \quad (B\infty)$$

over the set $P_2(\mathbb{R}^n)$. In this section, we consider existence, uniqueness, and regularity of the barycenter. We also prove generalizations for other distributions of measures and other underlying spaces.

3.1 Existence of the barycenter

Rather than proving the existence of a barycenter directly, we will, loosely speaking, approximate $(B\infty)$ by

$$\mu \mapsto \sum_{i=1}^N W_2^2(\mu_{\frac{i}{N}}, \mu) \quad (BN)$$

and take the limit as N tends to infinity. This approach will prove useful later when we establish the regularity of the barycenter.

Proposition 3.1.1. *A barycenter (a measure $\mu \in P_2(\mathbb{R}^n)$ minimizing $(B\infty)$) exists and it is supported on the convex hull of M .*

Proof. The result of Agueh and Carlier implies the existence of a minimizer μ^N for (BN) , as this is simply the barycenter for the measures $\mu_{\frac{i}{N}}$, $i = 1, \dots, N$, with equal weights. They also prove that it is supported on the set $\sum_{i=1}^N \frac{1}{N} M$, which is contained in the convex hull $c(M)$.

This yields, for all $\nu \in P_2(\mathbb{R}^n)$,

$$\sum_{i=1}^N W_2^2(\mu_{\frac{i}{N}}, \mu^N) \leq \sum_{i=1}^N W_2^2(\mu_{\frac{i}{N}}, \nu)$$

or,

$$\frac{1}{N} \sum_{i=1}^N W_2^2(\mu_{\frac{i}{N}}, \mu^N) \leq \frac{1}{N} \sum_{i=1}^N W_2^2(\mu_{\frac{i}{N}}, \nu)$$

Consider now the sequence μ^N ; by Prokhorov's theorem and the tightness of the set $P(c(M))$, we can assume, up to extraction of a subsequence, that μ^N converges weakly. This implies that μ^N converges in the Wasserstein metric, and so, letting $\mu^\infty \in P(c(M))$ be the weak limit, $W_2(\mu^N, \mu^\infty) \rightarrow 0$.

Now, by the triangle inequality, for any $\nu \in P_2(\mathbb{R}^n)$, we have

$$\begin{aligned} \frac{1}{N} \sum_{i=1}^N W_2^2(\mu_{\frac{i}{N}}, \mu^\infty) &\leq \frac{1}{N} \sum_{i=1}^N [W_2(\mu_{\frac{i}{N}}, \mu^N) + W_2(\mu^\infty, \mu^N)]^2 \\ &= \frac{1}{N} \sum_{i=1}^N W_2^2(\mu_{\frac{i}{N}}, \mu^N) + \frac{1}{N} \sum_{i=1}^N W_2^2(\mu^\infty, \mu^N) + 2 \frac{1}{N} \sum_{i=1}^N W_2(\mu^\infty, \mu^N) W_2(\mu_{\frac{i}{N}}, \mu^N) \\ &\leq \frac{1}{N} \sum_{i=1}^N W_2^2(\mu_{\frac{i}{N}}, \nu) + \frac{1}{N} \sum_{i=1}^N W_2^2(\mu^\infty, \mu^N) + 2W_2(\mu^\infty, \mu^N) \frac{1}{N} \sum_{i=1}^N W_2(\mu_{\frac{i}{N}}, \mu^N) \\ &= \frac{1}{N} \sum_{i=1}^N W_2^2(\mu_{\frac{i}{N}}, \nu) + W_2^2(\mu^\infty, \mu^N) + 2W_2(\mu^\infty, \mu^N) \frac{1}{N} \sum_{i=1}^N W_2(\mu_{\frac{i}{N}}, \mu^N) \end{aligned} \quad (1)$$

Now, as the compact set $c(M) \subseteq \mathbb{R}^n$ is bounded and $\mu_{\frac{i}{N}}$ and μ^N are supported on $c(M)$, we have, for some M , $|x - y|^2 < M$, whenever $x \in \text{spt}(\mu^N)$ and $y \in \text{spt}(\mu_{\frac{i}{N}})$. Therefore

$$W_2(\mu_{\frac{i}{N}}, \mu^N) \leq \sqrt{M}$$

and so the last term in inequality (1) is bounded above by

$$2W_2(\mu^\infty, \mu^N) \sqrt{M}.$$

Now, as $N \rightarrow \infty$, $\mu^N \rightarrow \mu^\infty$ in the Wasserstein metric, and so the last two terms above tend to 0. As the curve $t \mapsto \mu_t$ is continuous with respect to the Wasserstein distance, the mapping $t \mapsto W_2^2(\mu_t, \mu^\infty)$ is continuous, by the triangle inequality. Therefore, the quantity on the left hand side of inequality (1) tends to the Riemann integral of this curve as N tends to ∞ . A similar

conclusion holds for the first term on the right hand side, and so, taking the limit of as $N \rightarrow \infty$ in inequality (1) yields:

$$\int_0^1 W_2^2(\mu_t, \mu^\infty) dt \leq \int_0^1 W_2^2(\mu_t, \nu) dt$$

As this holds for any measure $\nu \in P_2(\mathbb{R}^n)$, this means that μ^∞ is a barycenter. \square

3.2 Uniqueness of the barycenter

In this section, we establish uniqueness of the barycenter, under Assumption A.

Lemma 3.2.1. *Fix $\nu \in P_2(\mathbb{R}^n)$. The function $P_2(\mathbb{R}^n) \ni \mu \mapsto W_2^2(\nu, \mu)$ is convex on $P_2(\mathbb{R}^n)$. If ν is absolutely continuous with respect to Lebesgue, it is strictly convex.*

Note that convexity here does *not* mean displacement convexity in the sense of McCann [20]; instead it means convexity with respect to the usual linear structure on the space of probability measures. This type of convexity is well known, and has been exploited in, for example, [16]. To the best of my knowledge, however, the strict convexity has not been explored.

Proof. Choose two measures μ_0 and μ_1 in $P_2(\mathbb{R}^n)$. For a fixed t , let γ_i be optimal couplings between μ_i and ν , for $i = 0, 1$, respectively. Now, let $\mu_s = s\mu_1 + (1-s)\mu_0$ and set $\gamma_s = s\gamma_1 + (1-s)\gamma_0$. Note that γ_s is a coupling of μ_s and ν . We then have

$$\begin{aligned} W_2^2(\nu, \mu_s) &\leq \int_{\mathbb{R}^n \times \mathbb{R}^n} |x - y|^2 d\gamma_s \\ &= s \int_{\mathbb{R}^n \times \mathbb{R}^n} |x - y|^2 d\gamma_1 + (1-s) \int_{\mathbb{R}^n \times \mathbb{R}^n} |x - y|^2 d\gamma_0 \\ &= sW_2^2(\nu, \mu_1) + (1-s)W_2^2(\nu, \mu_0) \end{aligned} \tag{2}$$

This establishes convexity of the function $\mu \mapsto W_2^2(\nu, \mu)$. We now show that this convexity is strict if ν is absolutely continuous with respect to Lebesgue measure. In this case, Brenier's theorem implies the existence of an optimal map $F_s : \text{spt}(\nu) \rightarrow \text{spt}(\mu_s)$ for each s , such that the unique optimal measure $\overline{\gamma_s}$ coupling ν and μ_s is concentrated on the graph $\{(x, F_s(x))\}$ [4].

Assume now that $\mu_0 \neq \mu_1$ and that $0 < s < 1$; we need to show that the inequality above is strict. Note first that the inequality is strict unless γ_s is an *optimal* coupling between ν and μ_s .

Now, as $\mu_0 \neq \mu_1$, the set $\{x : F_0(x) \neq F_1(x)\}$ has positive measure. Note that for each x where $F_0(x) \neq F_1(x)$, the coupling γ_s *splits* the mass at the point x between $F_0(x)$ and $F_1(x)$; for such x , both $(x, F_0(x))$ and $(x, F_1(x))$ belong to the support of the measure γ_s . On the other hand, the *optimal* measure $\overline{\gamma_s}$

coupling ν and μ_s is concentrated on the graph of a function F_s , and so, for ν almost all x , there is only one point (x, y) in the support of the optimizer (namely $(x, F_s(x))$). This immediately implies that γ_s is not the optimal coupling of ν and μ_s and so we must have a strict inequality. This completes the proof. \square

The preceding lemma easily implies the following result.

Lemma 3.2.2. *The function $\mu \mapsto \int_0^1 W_2^2(\mu_t, \mu) dt$ is convex on $P_2(\mathbb{R}^n)$. If Assumption A is satisfied, the function is strictly convex.*

Proof. Let $\mu_0, \mu_1 \in P_2(\mathbb{R}^n)$. The preceding lemma implies that for all $t \in [0, 1]$ and all $s \in (0, 1)$ we have

$$W_2^2(\mu_t, \mu_s) \leq s W_2^2(\mu_t, \mu_1) + (1 - s) W_2^2(\mu_t, \mu_0)$$

and the inequality is strict on a subset of $[0, 1]$ of positive measure. Integrating with respect to t yields the desired result. \square

This result immediately implies the uniqueness of the barycenter.

Corollary 3.2.3. *Under regularity Assumption A, the barycenter is unique.*

3.3 Regularity of the barycenter

In this subsection we obtain a regularity result on the barycenter μ^∞ , which will be crucial to our construction on the optimal stochastic process in section 4. Agueh and Carlier [1] proved the following regularity result for the barycenter of finitely many measures (ie, a minimizer of (BN)); assume that, for at least one $i \in \{1, 2, 3, \dots, N\}$, the measure $\mu_{\frac{i}{N}}$ is absolutely continuous with respect to Lebesgue measure, with an L^∞ density g_i . Then the barycenter μ^N is absolutely continuous with an L^∞ density g^N and

$$\|g^N\|_{L^\infty} \leq N \|g_i\|_{L^\infty}$$

Our general strategy in this subsection is to approximate μ^∞ by barycenters μ^N of finitely many measures, much like in subsection 3.1, and then deduce a regularity result (ie, a bound on the L^∞ norm of the density) from the regularity of the μ^N . Of course, the bound above tends to infinity as N tends to infinity and so to accomplish this goal we will need a refined regularity result on the barycenters of finitely many measures, with a bound on the L^∞ norm which is uniform in N .

Proposition 3.3.1. *Let $\mu_i \in P_2(\mathbb{R}^n)$ for $i = 1, 2, \dots, N$ and suppose μ minimizes $\mu \mapsto \sum_{i=1}^N \lambda_i W_2^2(\mu, \mu_i)$ on $P_2(\mathbb{R}^n)$, where $0 < \lambda_i < 1$ and $\sum_{i=1}^\infty \lambda_i = 1$. Let $B \subseteq \{1, 2, \dots, N\}$ be nonempty and assume that for all $i \in B$, μ_i is absolutely continuous with respect to Lebesgue measure with an L^∞ density g_i . Then μ is absolutely continuous with respect to Lebesgue measure with an L^∞ density g satisfying:*

$$\|g\|_{L^\infty} \leq \left[\sum_{i \in B} \frac{\lambda_i}{\|g_i\|_{L^\infty}^{\frac{1}{n}}} \right]^{-n}$$

Proof. By a result of Agueh and Carlier, ([1], Proposition 3.8), for almost all x , $\sum_{i=1}^N \lambda_i Du_i(x) = x$, where Du_i is the Brenier map pushing the Barycenter μ^N forward to μ_i . As each convex function u_i is twice differentiable almost everywhere, we can differentiate this equation to obtain

$$\sum_{i=1}^N \lambda_i D^2 u_i(x) = I_n$$

for almost all x , where I_n is the $n \times n$ identity matrix. Taking determinants and n -th roots yields:

$$[\det \sum_{i=1}^N \lambda_i D^2 u_i(x)]^{\frac{1}{n}} = 1$$

As each $D^2 u_i(x)$ is symmetric and positive definite wherever it exists, Minkowski's determinant inequality combined with the preceding equation yields:

$$\sum_{i=1}^N \lambda_i (\det D^2 u_i(x))^{\frac{1}{n}} \leq [\det \sum_{i=1}^N \lambda_i D^2 u_i(x)]^{\frac{1}{n}} = 1.$$

As each term $\lambda_i (\det D^2 u_i(x))^{\frac{1}{n}}$ is non-negative, we obtain

$$\sum_{i \in B} \lambda_i (\det D^2 u_i(x))^{\frac{1}{n}} \leq \sum_{i=1}^N \lambda_i (\det D^2 u_i(x))^{\frac{1}{n}} \leq 1.$$

Now, using the result of Agueh and Carlier, we know that μ is absolutely continuous, ie $d\mu = g(x)dx$, and it is well known that for each $i \in B$, u_i solves the Monge-Ampere equation almost everywhere, $\det D^2 u_i(x) = \frac{g(x)}{g_i(Du_i(x))}$. Combined with the preceding inequality, this implies

$$[g(x)]^{\frac{1}{n}} \leq \left[\sum_{i \in B} \frac{\lambda_i}{g_i(Du_i(x))^{\frac{1}{n}}} \right]^{-1} \leq \left[\sum_{i \in B} \frac{\lambda_i}{\|g_i\|_{L^\infty}^{\frac{1}{n}}} \right]^{-1}$$

□

Lemma 3.3.2. *Assume Assumption B and let $m_K > 0$ be the Lebesgue measure of the set A_K . Then there exists a sequence of measures $\bar{\mu}^N$, absolutely continuous with respect to Lebesgue measure, with densities $\bar{g}^N(x)$ satisfying $\|\bar{g}^N\|_{L^\infty} \leq \frac{K}{m_K^n}$, converging weakly to μ^∞ .*

Proof. For $i = 0, 1, \dots, N-1$, set $I_i = [\frac{i}{N}, \frac{i+1}{N}]$. Let $B_K = \{i : I_i \cap A_K \neq \emptyset\}$ be subset of indices i for which I_i contains at least one point in the set A_K . The union $\bigcup_{i \in B_K} I_i$ clearly covers A_K , and so, denoting the size of B_K by $|B_K|$, we must have

$$\frac{|B_K|}{N} \geq m_K.$$

Now, we choose $t_i \in I_i$ and approximate the Riemann integral much like in the proof of existence, except that, whenever i is in B_K , we choose the point $t_i \in I_i \cap A_K$, rather than taking $t_i = \frac{i}{N}$. We define $\bar{\mu}^N$ to be the barycenter of the measures μ_{t_i} , with equal weights $\lambda_i = \frac{1}{N}$; that is the minimizer on $P_2(\mathbb{R}^n)$ of:

$$\mu \mapsto \frac{1}{N} \sum_{i=1}^N W_2^2(\mu_{t_i}, \mu)$$

Our result above implies that the barycenter $\bar{\mu}^N$ is absolutely continuous with respect to Lebesgue measure, with a density $\bar{g}^N(x)$ satisfying

$$\begin{aligned} \|\bar{g}^N\|_{L^\infty} &\leq \left[\sum_{i \in B_K} \frac{\lambda_i}{\|g_{t_i}\|_{L^\infty}^{\frac{1}{n}}} \right]^{-n} \\ &\leq \left[\sum_{i \in B_K} \frac{1}{NK^{\frac{1}{n}}} \right]^{-n} \\ &= \left[\frac{|B_K|}{NK^{\frac{1}{n}}} \right]^{-n} \\ &= \frac{N^n K}{|B_K|^n} \\ &\leq \frac{K}{m_K^n} \end{aligned}$$

Now, up to extraction of a subsequence, $\bar{\mu}^N$, converges weakly by Prokhorov's theorem to some measure $\bar{\mu}^\infty$. Exactly as in the proof of existence, one can prove that $\bar{\mu}^\infty$ is a barycenter. It then follows by the uniqueness result in the last subsection that $\bar{\mu}^\infty = \mu^\infty$. \square

By approximation, we then easily obtain the following regularity result on our barycenter μ^∞ .

Corollary 3.3.3. *Under Assumption B, the barycenter is absolutely continuous with respect to Lebesgue measure and its density $g^\infty(x)$ satisfies $\|g^\infty\|_{L^\infty} \leq \frac{K}{m_K^n}$.*

Proof. It suffices to prove $\mu^\infty(A) \leq \frac{K}{m_K} |A|$ for any Borel set $A \subseteq c(M)$. If A is open, this follows easily from Lemma 3.3.1, as the weak convergence $\bar{\mu}^N \rightarrow \mu^\infty$ implies

$$\mu^\infty(A) \leq \liminf \bar{\mu}^N(A) \leq \frac{K}{m_K}|A|.$$

If A is not open, we may, for any $\epsilon > 0$, find an open set U such that $A \subseteq U$ and $|U \setminus A| \leq \epsilon$. Then we have

$$\begin{aligned} \mu^\infty(A) &\leq \mu^\infty(U) \\ &\leq \frac{K}{m_K}|U| \\ &= \frac{K}{m_K}(|A| + |U \setminus A|) \\ &\leq \frac{K}{m_K}|A| + \frac{K}{m_K}\epsilon \end{aligned}$$

Taking the limit as $\epsilon \rightarrow 0$ yields the desired result. \square

3.4 Generalization to other spaces and distributions

The purpose of this subsection is to demonstrate that our approach to uniqueness of the barycenter holds for more general underlying spaces M and more general distributions of measures. The results of this subsection are not essential to the rest of the paper and can safely be skipped.

For this subsection only, let (M, g) be a compact Riemannian manifold and $P(M)$ denote the set of Borel probability measures on M . Given probability measures μ and ν in $P(M)$, the Wasserstein distance between μ and ν is defined as in Euclidean space, with the Riemannian distance squared replacing the Euclidean distance:

$$W_2^2(\mu, \nu) = \inf \int_{M \times M} d^2(x, y) d\gamma(x, y),$$

where the infimum is over all measures γ on $M \times M$ projecting to μ and ν , respectively.

Now, let Γ be a probability measure on $P(M)$. A barycenter of Γ is a minimizer of

$$\mu \mapsto \int W_2^2(\mu, \nu) d\Gamma(\nu).$$

By continuity and Prokhorov's theorem, it is straightforward to verify that a barycenter exists; see, for example, [21]. We note here that our proof of uniqueness relied only on existence and uniqueness of Monge solutions for arbitrary μ and a set of ν of positive Γ measure. Let $P_{ac}(M)$ be the set of Borel probability measures on M which are absolutely continuous with respect to local coordinates. By McCann's theorem [19], whenever $\nu \in P_{ac}(M)$, there is a unique optimal map between ν and μ . Therefore, we obtain:

Theorem 3.4.1. *Suppose that $P_{ac}(M) \subseteq P(M)$ has positive Γ measure. Then the barycenter of γ is unique.*

Assuming M is a bounded subset of \mathbb{R}^n , when Γ has finite support, this yields the uniqueness theorem of Agueh and Carlier. When Γ is supported on a Wasserstein continuous curve, we recover our results from a previous section.

4 Infinitely many marginals

4.1 Construction and basic properties of the optimal process

We now return to our problem of primary present interest, namely the optimal transportation problem with infinitely many marginals, (MK_∞) .

We will use the barycenter from the previous section to construct a stochastic process X_t^{opt} . We will then show that this process is the unique minimizer in (MK_∞) .

We construct our optimal process X_t^{opt} as follows.

Definition 4.1.1. *We take our underlying probability space to be \mathbb{R}^n , with the barycenter μ^∞ . Then taking Du_t to be the Brenier map pushing μ^∞ forward to μ_t , we define a stochastic process by*

$$X_t^{opt}(x) = Du_t(x),$$

with the barycenter μ^∞ as the underlying probability space.

Note that this definition means that the sample paths of the optimal process X_t^{opt} are $t \mapsto Du_t(x)$, for $x \in c(M)$, with a probability given by μ^∞ . Recall that a stochastic process Y_t is continuous in probability (or in measure) if, for all $t \in [0, 1]$ and all $\epsilon > 0$.

$$\lim_{s \rightarrow t} \mathbb{P}(|Y_s - Y_t| > \epsilon) = 0$$

To prove measurability of X_t^{opt} , we will need the following proposition.

Proposition 4.1.2. *The process X_t^{opt} is continuous in probability.*

Proof. Recall that the path μ_t is weakly continuous; this then easily follows from a well known result on the stability of optimal transportation (see, for example, Villani [29], Corollary 5.23). \square

Recall that a stochastic process Y_t is a *version* of another process Z_t if, for all $t \in [0, 1]$, $Y_t = Z_t$ almost surely. In this case, we say that Y_t and Z_t are *stochastically equivalent*. It is a well known result that every stochastic process which is continuous in probability has a measurable version, and so the preceding proposition implies:

Corollary 4.1.3. *The process X_t^{opt} has a measurable version.*

In light of the preceding corollary, we assume from now on that the process X_t^{opt} is measurable.

4.2 Proof of optimality

Our aim is now to prove that the process X_t^{opt} defined in the last subsection is in fact optimal for (MK_∞) . As a preliminary step in this direction, we will need to show that the *average measure*, defined by

$$\mu_a^{opt} = \text{law}\left(\int_0^1 X_t^{opt} dt\right)$$

is in fact the barycenter μ^∞ .

Proposition 4.2.1. *Assume Assumption B. Then, for the process X_t^{opt} from Definition 4.1.1, the average measure coincides with the barycenter: $\mu_a^{opt} = \mu^\infty$.*

Proof. By Corollary 3.3.3, the barycenter μ^∞ is absolutely continuous with respect to Lebesgue measure. Now, for each t , the function

$$\mu \rightarrow W_2^2(\mu_t, \mu),$$

restricted to the set $P_{ac,2}(\mathbb{R}^n)$ of absolutely continuous measures with finite second moments, is differentiable with respect to the Wasserstein structure on $P_{ac,2}(\mathbb{R}^n)$ [2]. This means that given a curve μ_s in $P_{ac,2}(\mathbb{R}^n)$ with $\mu_0 = \mu^\infty$, we have:

$$\left.\frac{d}{ds}\right|_{s=0} W_2^2(\mu_s, \mu_t) = 2 \int_{\mathbb{R}^n} \langle y - Du_t(y), \xi_0(y) \rangle d\mu^\infty(y)$$

where $\xi_s(y)$ is a vector field satisfying $\frac{\partial \mu_s}{\partial s} + D \cdot (\mu_s \xi_s(y)) = 0$; that is, the tangent to μ_s in $P_{ac,2}(\mathbb{R}^n)$. Note that we are abusing notation slightly by identifying the measure μ_s with its density.

Using the dominated convergence theorem, this means that

$$\mu \mapsto \int_0^1 W_2^2(\mu_t, \mu) dt$$

is differentiable on $P_{ac,2}(\mathbb{R}^n)$ as well, and so its derivative must vanish at the minimizer, μ^∞ . Using the formula for the derivative, we have,

$$0 = \int_0^1 \int_{\mathbb{R}^n} \langle y - Du_t(y), \xi_0(y) \rangle d\mu^\infty(y) dt$$

for any tangent vector field ξ_0

By Fubini's theorem,

$$0 = \int_{\mathbb{R}^n} \left\langle \left(\int_0^1 y - Du_t(y) dt \right), \xi_0(y) \right\rangle d\mu^\infty(y) \quad (3)$$

Note this holds for all tangent vector fields ξ_0 to $P_{ac,2}(\mathbb{R}^n)$ at μ^∞ . As each u_t is a convex function, the integral $v(x) = \int_0^1 u_t(x) dt$ is also convex, and again using the dominated convergence theorem we have:

$$Dv(x) = \int_0^1 Du_t(x) dt$$

In particular, we can take $\xi_0(y) = y - \int_0^1 Du_t(y)dt = \int_0^1 (y - Du_t(y))dt$ in (3), to obtain

$$0 = \int_{\mathbb{R}^n} \left\langle \left(\int_0^1 y - Du_t(y)dt \right), \left(\int_0^1 y - Du_t(y)dt \right) \right\rangle d\mu^\infty(y) = \int_{\mathbb{R}^n} \left| \left(\int_0^1 y - Du_t(y)dt \right) \right|^2 d\mu^\infty(y)$$

This implies

$$0 = \int_0^1 y - Du_t(y)dt$$

μ^∞ almost everywhere. Therefore, $y \mapsto \int_0^1 Du_t(y)dt = \int_0^1 X_t^{opt}(y)dt$ is the identity mapping. As this map pushes μ^∞ forward to μ_a , this immediately implies the desired result. \square

Theorem 4.2.2. *Assume Assumption B. Then the X_t^{opt} from Definition 4.1.1 is optimal for (MK_∞) . It is the unique optimizer in the sense that, if Y_t is any other optimal process, we have for almost all t , $X_t^{opt} = Y_t$, almost surely.*

Proof. It is clear from the construction that $\text{law} X_t^{opt} = \mu_t$ for all t . Now, take any stochastic process Y_t , such that $\text{law}(Y_t) = \mu_t$. Denote by μ_a the law of the random variable $Y_a = \int_0^1 Y_t dt$. We will denote by $\mu_{t,a}$ the law of the ordered pair (Y_t, Y_a) on $\mathbb{R}^n \times \mathbb{R}^n$; note that this implies that μ_t and μ_a are the marginals of $\mu_{t,a}$. Now, note that:

$$\begin{aligned} E\left(\int_0^1 (|Y_t - \int_0^1 Y_s ds|^2) dt\right) &= E\left(\int_0^1 |Y_t|^2 dt - 2 \int_0^1 Y_t dt \int_0^1 Y_s ds + \left|\int_0^1 Y_s ds\right|^2\right) \\ &= E\left(\int_0^1 |Y_t|^2 dt - \left|\int_0^1 Y_s ds\right|^2\right) \\ &= \int_0^1 E|Y_t|^2 dt - E\left|\int_0^1 Y_s ds\right|^2 \\ &= \int_0^1 \int_{\mathbb{R}^n} |Y_t|^2 d\mu_t dt - E\left|\int_0^1 Y_s ds\right|^2 \end{aligned}$$

Note that the first term above depends only on $\text{law}(Y_t) = \mu_t$. Therefore, maximizing $E\left(\left|\int_0^1 Y_s ds\right|^2\right)$ subject to the constraint $\text{law}(Y_t) = \mu_t$ is equivalent to minimizing $E\left(\int_0^1 (|Y_t - \int_0^1 Y_s ds|^2) dt\right)$, subject to the same constraint.

Now, we have, using Fubini,

$$\begin{aligned}
E\left(\int_0^1 (|Y_t - \int_0^1 Y_s ds|) dt\right) &= \int_0^1 \left(E(|Y_t - \int_0^1 Y_s ds|^2)\right) dt \\
&= \int_0^1 \left(\int_{\mathbb{R}^n \times \mathbb{R}^n} |Y_t - Y_a|^2 d\mu_{t,a}\right) dt \\
&\geq \int_0^1 W_2^2(\mu_t, \mu_a) dt \\
&\geq \int_0^1 W_2^2(\mu_t, \mu^\infty) dt
\end{aligned}$$

Observe that we have equality if and only if

1. μ_a is the barycenter μ^∞ of the μ_t 's and,
2. for almost all t , the measure $\mu_{t,a}$ is the optimal coupling between μ_t and μ_a .

Now, assuming the first condition, $\mu_a = \mu^\infty$ is absolutely continuous with respect to Lebesgue measure by Corollary 3.3.3, and so the optimal coupling $\mu_{t,a}$ is concentrated on the graph of the function $x \mapsto Du_t(x)$. Therefore, these two conditions imply that the sample path Y_t is completely determined almost surely by Y_a , which is distributed according to the barycenter. We can therefore take the underlying probability space to be μ^∞ and the second condition implies that the process $Y_t = X_t^{opt}$ almost surely, for almost all t . \square

We can obtain a more elegant uniqueness result if we restrict our attention to stochastic processes which are continuous in probability; the following theorem implies that X_t^{opt} is the unique, continuous in probability maximizer for (MK_∞) , modulo stochastic equivalence.

Theorem 4.2.3. *Assume Assumption B and suppose Y_t is optimal for (MK_∞) and Y_t is continuous in probability. Then Y_t is a version of X_t^{opt} .*

Proof. From our previous uniqueness result, we know that $Y_t = X_t^{opt}$ almost surely, for almost all t . We need to prove this for *all* t .

Fix $t_0 \in [0, 1]$. Then we can choose a sequence t_i converging to t_0 such that $Y_{t_i} = X_{t_i}$ almost surely. By continuity in probability, Y_{t_i} converges to Y_{t_0} in probability and X_{t_i} converges to X_{t_0} in probability. This immediately implies $X_{t_0} = Y_{t_0}$ almost surely, as desired. \square

We now prove an analogue of Brenier's theorem [4] (for two marginal problems) and the result of Gangbo and Swiech [14] (for several marginals). In our context, it is natural to interpret this result as saying that we can take the underlying probability space of our stochastic process to be $M \subseteq \mathbb{R}^n$.

Theorem 4.2.4. (*Monge solutions*) Assume Assumption B holds and suppose $\mu_{t_0} \in P_{ac}(M)$. Then we can take μ_{t_0} to be the underlying probability space of the unique optimal process X_t^{opt} . That is, the optimal process can be written as $X_t^{opt} = F_t(X_{t_0}^{opt})$, where, for each t , $F_t : \mathbb{R}^n \rightarrow \mathbb{R}^n$ is a mapping pushing μ_{t_0} forward to μ_t , and F_{t_0} is the identity mappings.

Proof. As μ_{t_0} does not charge small sets, Brenier's theorem implies that the optimal map Du_{t_0} pushing the barycenter μ^∞ forward to μ_{t_0} is invertible almost everywhere; its inverse is $Du_{t_0}^*$, where $u_{t_0}^*$ is the Legendre transform of u . We then have $X_t^{opt}(x) = Du_t(x) = Du_t(Du_{t_0}^*(z))$, where x is distributed according to μ^∞ and $z = Du_{t_0}(x)$ is distributed according to μ_{t_0} . Taking $F_t = Du_t \circ Du_{t_0}^*$ yields the desired result. \square

This result means that the stochastic process X_t^{opt} is *deterministic*, in the sense that if we know $X_{t_0}^{opt}$, we know X_t^{opt} for all t .

References

- [1] M. Agueh and G. Carlier. Barycenters in the Wasserstein space. *SIAM J. Math. Anal.*, 43(2):904–924, 2011.
- [2] L. Ambrosio, N. Gigli, and G. Savaré. *Gradient flows in metric spaces and in the space of probability measures*. Lectures in Mathematics, ETH Zürich. Birkhäuser Verlag, Basel, 2005.
- [3] M. Bernot, J. Delon, G. Peyre, and J. Rabin. Wasserstein barycenter and its application to texture mixing. Preprint available at Hal-00476064.
- [4] Y. Brenier. Decomposition polaire et rearrangement monotone des champs de vecteurs. *C.R. Acad. Sci. Paris. Ser. I Math.*, 305:805–808, 1987.
- [5] L. Caffarelli. Allocation maps with general cost functions. In *Partial Differential Equations and Applications*, volume 177 of *Lecture Notes in Pure and Applied Math*, pages 29–35. Dekker, New York, 1996.
- [6] G. Carlier. On a class of multidimensional optimal transportation problems. *J. Convex Anal.*, 10(2):517–529, 2003.
- [7] G. Carlier and I. Ekeland. Matching for teams. *Econom. Theory*, 42(2):397–418, 2010.
- [8] G. Carlier and B. Nazaret. Optimal transportation for the determinant. *ESAIM Control Optim. Calc. Var.*, 14(4):678–698, 2008.
- [9] P-A. Chiapporri, R. McCann, and L. Nesheim. Hedonic price equilibria, stable matching and optimal transport; equivalence, topology and uniqueness. *Econom. Theory*, 42(2):317–354, 2010.
- [10] C. Cotar, G. Friesecke, and C. Klüppelberg. Preprint.

- [11] C. Cotar, G. Friesecke, and C. Klüppelberg. Density functional theory and optimal transportation with Coulomb cost. Preprint available at arXiv:1104.0603.
- [12] W. Gangbo. PhD thesis, Universite de Metz, 1995.
- [13] W. Gangbo and R. McCann. The geometry of optimal transportation. *Acta Math.*, 177:113–161, 1996.
- [14] W. Gangbo and A. Świąch. Optimal maps for the multidimensional monge-kantorovich problem. *Comm. Pure Appl. Math.*, 51(1):23–45, 1998.
- [15] H. Heinich. Probleme de Monge pour n probabilités. *C.R. Math. Acad. Sci. Paris*, 334(9):793–795, 2002.
- [16] Richard Jordan, David Kinderlehrer, and Felix Otto. The variational formulation of the fokker-planck equation. *SIAM Journal on Mathematical Analysis*, 29(1):1–17, 1998.
- [17] M. Knott and C. Smith. On a generalization of cyclic monotonicity and distances among random vectors. *Linear Algebra Appl.*, 199:363–371, 1994.
- [18] V. Levin. Abstract cyclical monotonicity and Monge solutions for the general Monge-Kantorovich problem. *Set-Valued Analysis*, 7(1):7–32, 1999.
- [19] R. McCann. Polar factorization of maps on Riemannian manifolds. *Geom. Funct. Anal.*, 11:589–608, 2001.
- [20] R.J. McCann. A convexity theory for interacting gases and equilibrium crystals. *Adv. Math.*, 128(1):153–179, 1997.
- [21] S. Ohta. Barycenters in Alexandrov spaces of curvature bounded below. Preprint available at <http://www.math.kyoto-u.ac.jp/~sohta/papers/bary.pdf>. To appear in *Adv. Geom.*
- [22] I. Olkin and S.T. Rachev. Maximum submatrix traces for positive definite matrices. *SIAM J. Matrix Ana. Appl.*, 14:390–39, 1993.
- [23] B. Pass. On a class of optimal transportation problems with infinitely many marginals. Preprint.
- [24] B. Pass. Uniqueness and monge solutions in the multimarginal optimal transportation problem. *SIAM Journal on Mathematical Analysis*, 43(6):2758–2775, 2011.
- [25] B. Pass. On the local structure of optimal measures in the multi-marginal optimal transportation problem. *Calculus of Variations and Partial Differential Equations*, 43:529–536, 2012. 10.1007/s00526-011-0421-z.
- [26] L. Rüschendorf and L. Uckelmann. On optimal multivariate couplings. In *Proceedings of Prague 1996 conference on marginal problems*, pages 261–274. Kluwer Acad. Publ., 1997.

- [27] L. Rüschendorf and L. Uckelmann. On the n -coupling problem. *J. Multivariate Anal.*, 81:242–258, 2002.
- [28] K.T. Sturm. Probability measures on metric spaces of nonpositive curvature. *Contemp. Math.*, 338:357–390, 2003.
- [29] C. Villani. *Optimal transport: old and new*, volume 338 of *Grundlehren der mathematischen Wissenschaften*. Springer, New York, 2009.